Using Big Data to Solve Economic and Social Problems

Professor Raj Chetty
Head Section Leader: Gregory Bruich, Ph.D.
Harvard University

Spring 2019
Improving Health Outcomes

- Research in economics typically focuses on earnings or wealth as key outcomes of interest

- But most people view health and life expectancy as among the most important aspects of well-being

- What interventions are most effective in improving health (holding fixed current frontier of medical technology)?
  - Research on these issues spans multiple fields, from epidemiology and public health to economics
Improving Health Outcomes: Overview

- This part of the class illustrates how big data is helping us learn how to improve health, in three segments:

  1. Descriptive analysis of health outcomes in U.S. population
     [method: survival analysis]

Improving Health Outcomes: Overview

- This part of the class illustrates how big data is helping us learn how to improve health, in three segments:

1. Descriptive analysis of health outcomes in U.S. population [method: survival analysis]

2. Economics applications: impacts of food stamps (Jesse Shapiro) and health insurance [method: regression discontinuities]

Improving Health Outcomes: Overview

- This part of the class illustrates how big data is helping us learn how to improve health, in three segments:

1. Descriptive analysis of health outcomes in U.S. population [method: survival analysis]

2. Economics applications: impacts of food stamps (Jesse Shapiro) and health insurance [method: regression discontinuities]

3. Epidemiology application: using big data to forecast pandemics [method: predictive modeling]


Income and Life Expectancy

- Most common measure of health: mortality rates
  - Crude but well measured in population data

- Begin with basic descriptive facts about life expectancy in America

- Chetty et al. (2016) examine relationship between life expectancy and income
  - Use data on entire U.S. population from 1999-2013 (1.4 billion observations)
Estimating Life Expectancy: Data

- Mortality measured using Social Security death records

- Income measured at household level using tax returns

- Focus on percentile ranks in income distribution
  - Rank individuals in national income distribution within birth cohort, gender, and tax year
Methodology to Estimate Life Expectancy

- **Goal**: estimate expected age of death conditional on an individual’s income at age 40, controlling for differences in race and ethnicity

  - *Period* life expectancy: life expectancy for a hypothetical individual who experiences mortality rates at each age observed in a given year

- **Three steps**:
  1. Calculate mortality rates by income rank and age for observed ages
  2. Estimate a survival model to extrapolate to older ages
  3. Adjust for racial differences in mortality rates
Survival Curves for Men at 5th and 95th Percentiles

Age 76
Survival Curves for Men at 5th and 95th Percentiles

- p5 Survival Rate: 52%
- p95 Survival Rate: 83%

Age 76
Step 2: Predicting Mortality Rates at Older Ages

- To calculate life expectancy, need estimates of mortality rates beyond age 76

- Gompertz (1825) documented a robust empirical pattern: mortality rates grow exponentially with age

\[ m(a) = k \ e^{\beta a} \]

\[ \Rightarrow \log m(a) = \kappa + \beta a \]
Log Mortality Rates for Men at 5th and 95th Percentiles

Gompertz: p95

Data: p5

Gompertz: p5

Data: p95

Log Mortality Rate

Age in Years
Log Mortality Rates for Men at 5th and 95th Percentiles

Medicare Eligibility Threshold

Age 65

Log Mortality Rate

Age in Years

Data: p5  Gompertz: p5  Data: p95  Gompertz: p95
Survival Curves for Men at 5th and 95th Percentiles

- Data: p5
- Gompertz: p5
- Data: p95
- Gompertz: p95

NCHS and SSA Estimates (constant across income groups)

Survival Rate (%) vs. Age in Years (a)
National Statistics on Income and Life Expectancy
Expected Age at Death vs. Household Income Percentile
For Men at Age 40
Expected Age at Death vs. Household Income Percentile
For Men at Age 40

Bottom 1%: 72.7 Years
Top 1%: 87.3 Years
Expected Age at Death vs. Household Income Percentile
By Gender at Age 40

Women, Bottom 1%: 78.8
Women, Top 1%: 88.9
Men, Bottom 1%: 72.7
Men, Top 1%: 87.3
Expected Age at Death vs. Household Income Percentile
By Gender at Age 40

- Women, Bottom 1%: 78.8 years
- Women, Top 1%: 88.9 years
- Men, Bottom 1%: 72.7 years
- Men, Top 1%: 87.3 years

- Bottom 1% Gender Gap: 6.1 years
- Top 1% Gender Gap: 1.6 years
Time Trends

- How are gaps in life expectancy changing over time?
Trends in Expected Age at Death by Income Quartile in the US
For Men Age 40, 2001-2014

- 4th Quartile: Annual Change = 0.20 (0.17, 0.24)
- 3rd Quartile: Annual Change = 0.18 (0.15, 0.20)
- 2nd Quartile: Annual Change = 0.12 (0.08, 0.16)
- 1st Quartile: Annual Change = 0.08 (0.05, 0.11)
Trends in Expected Age at Death by Income Quartile in the US
For Women Age 40, 2001-2014

- 1st Quartile: Annual Change = 0.10 (0.06, 0.13)
- 2nd Quartile: Annual Change = 0.17 (0.13, 0.20)
- 3rd Quartile: Annual Change = 0.25 (0.22, 0.28)
- 4th Quartile: Annual Change = 0.23 (0.20, 0.25)
Local Area Variation in Life Expectancy by Income
Expected Age at Death vs. Household Income for Men in Selected Cities
Expected Age at Death vs. Household Income for Women in Selected Cities

Household Income Percentile vs. Expected Age at Death

- New York City
- San Francisco
- Dallas
- Detroit
Expected Age at Death for 40 Year Old Men
Bottom Quartile of U.S. Income Distribution

Note: Lighter Colors Represent Areas with Higher Life Expectancy
Expected Age at Death for 40 Year Old Men
Pooling All Income Groups

Note: Lighter Colors Represent Areas with Higher Life Expectancy
Expected Age at Death for 40 Year Old Women
Bottom Quartile of U.S. Income Distribution

Note: Lighter Colors Represent Areas with Higher Life Expectancy
### Expected Age at Death for 40 Year Olds in Bottom Quartile
#### Top 10 and Bottom 10 CZs Among 100 Largest CZs

<table>
<thead>
<tr>
<th>Rank</th>
<th>Top 10 CZs</th>
<th>Expected Age at Death</th>
<th>Rank</th>
<th>Bottom 10 CZs</th>
<th>Expected Age at Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New York, NY</td>
<td>81.8 (81.6, 82.0)</td>
<td>91</td>
<td>San Antonio, TX</td>
<td>78.0 (77.6, 78.4)</td>
</tr>
<tr>
<td>2</td>
<td>Santa Barbara, CA</td>
<td>81.7 (81.3, 82.1)</td>
<td>92</td>
<td>Louisville, KY</td>
<td>77.9 (77.7, 78.2)</td>
</tr>
<tr>
<td>3</td>
<td>San Jose, CA</td>
<td>81.6 (81.2, 82.0)</td>
<td>93</td>
<td>Toledo, OH</td>
<td>77.9 (77.6, 78.2)</td>
</tr>
<tr>
<td>4</td>
<td>Miami, FL</td>
<td>81.2 (80.9, 81.6)</td>
<td>94</td>
<td>Cincinnati, OH</td>
<td>77.9 (77.7, 78.1)</td>
</tr>
<tr>
<td>5</td>
<td>Los Angeles, CA</td>
<td>81.1 (80.9, 81.4)</td>
<td>95</td>
<td>Detroit, MI</td>
<td>77.7 (77.5, 77.8)</td>
</tr>
<tr>
<td>6</td>
<td>San Diego, CA</td>
<td>81.1 (80.8, 81.4)</td>
<td>96</td>
<td>Tulsa, OK</td>
<td>77.6 (77.4, 77.9)</td>
</tr>
<tr>
<td>7</td>
<td>San Francisco, CA</td>
<td>80.9 (80.6, 81.3)</td>
<td>97</td>
<td>Indianapolis, IN</td>
<td>77.6 (77.4, 77.8)</td>
</tr>
<tr>
<td>8</td>
<td>Santa Rosa, CA</td>
<td>80.8 (80.5, 81.2)</td>
<td>98</td>
<td>Oklahoma City, OK</td>
<td>77.6 (77.3, 77.8)</td>
</tr>
<tr>
<td>9</td>
<td>Newark, NJ</td>
<td>80.7 (80.5, 80.9)</td>
<td>99</td>
<td>Las Vegas, NV</td>
<td>77.6 (77.4, 77.8)</td>
</tr>
<tr>
<td>10</td>
<td>Port St. Lucie, FL</td>
<td>80.7 (80.5, 80.9)</td>
<td>100</td>
<td>Gary, IN</td>
<td>77.4 (77.1, 77.8)</td>
</tr>
</tbody>
</table>

Note: 95% confidence intervals shown in parentheses
Why Does Life Expectancy for Low-Income Individuals Vary Across Areas?
Why Does Life Expectancy for Low-Income Individuals Vary Across Areas?

- Now use local area variation to explore determinants of life expectancy

- Key question: is lower life expectancy in some areas driven by lack of access to health care or differences in health behavior?

- Correlate life expectancy estimates with measure of health care access and health behaviors to answer this question
Correlations of Expected Age at Death with Health and Social Factors For Individuals in Bottom Quartile of Income Distribution
Smoking Rates for Individuals in Bottom Income Quartile

Note: Lighter Colors Represent Areas Lower Smoking Rates
Correlations of Expected Age at Death with Health and Social Factors
For Individuals in Bottom Quartile of Income Distribution
Why Does Life Expectancy for Low-Income Individuals Vary Across Areas?

- Local area variation suggests that differences in health behaviors are more predictive of life expectancy than differences in health care access.

- Further evidence for this view comes from directly examining nutritional patterns.
Differences in Nutrition by Income

- Alcott et al. (2018) use Nielsen homescan data on grocery store purchases to examine how nutrition varies with income
  - About 170,000 households who scan all of their purchases and record UPCs, which are then matched to nutritional information from the USDA
Healthfulness of Grocery Purchases by Household Income

Source: Allcott, Diamond, Dube, Handbury, Rahkovsky, and Schnell 2018
Healthfulness of Grocery Purchases by Household Income

Source: Allcott, Diamond, Dube, Handbury, Rahkovsky, and Schnell 2018
These differences in nutrition are not driven by a lack of access to health food ("food deserts")
Healthfulness of Grocery Purchases by Household Income that Shop in the Same Market

Within Zip Code x Year

Household Size and Age Controls

All Demographic Controls

Within Chain x Year

Household Size and Age Controls

All Demographic Controls

Source: Allcott, Diamond, Dube, Handbury, Rahkovsky, and Schnell 2018
Differences in Health Behaviors by Income

- These differences in nutrition are *not* driven by a lack of access to health food ("food deserts")

- Again suggests that differences in health outcomes are not caused by a direct lack of access to resources
  - Instead, appear to be due to different *choices* made by lower-income households
Differences in Health Behaviors by Income

- Why do low income households tend to have less healthy behaviors?

- One hypothesis: effects of environment and resources at early ages on preferences
  - Ex: Atkin (2016) studies migrants in India and shows that nutritional habits formed at young ages persist for many years after people move

- Alternative hypothesis: lack of income constrains choice (unhealthy foods may be less expensive per calorie)
  - Discuss this economic explanation in next lecture with Jesse Shapiro